

SPECIATION EFFECT IN THE PENNA AGING MODEL

A. ŁASZKIEWICZ, SZ. SZYMCZAK, and S. CEBRAT*

*Department of Genetics, Institute of Microbiology
University of Wrocław, ul. Przybyszewskiego 63/77
PL-54148 Wrocław, Poland
cebrat@microb.uni.wroc.pl

Received 26 November 2002
Revised 27 November 2002

We have simulated the evolution of diploid, sexually reproducing populations using the Penna model of aging. We have noted that diminishing the recombination frequency during the gamete production generates a specific diversity of genomes in the populations. When two populations independently evolving for some time were mixed in one environmental niche of the limited size and crossbreeding between them was allowed, the average lifespan of hybrids was significantly shorter than the lifespan of the individuals of parental lines. Another effect of higher hybrid mortality is the faster elimination of one parental line from the shared environment. The two populations living in one environment co-exist much longer if they are genetically separated — they compete as two species instead of crossbreeding. This effect can be considered as the first step to speciation — any barrier eliminating crossbreeding between these populations, leading to speciation, would favor the populations.

Keywords: Biological aging; Monte Carlo simulation; recombination; genetic linkage; speciation.

1. Introduction

One of the most important mechanisms speeding the evolution processes is genetic recombination, which could be understood as a reshuffling of information inside the genomes as well as exchange of information between genomes. Nevertheless, there is a strong barrier in the exchange of information between organisms belonging to different species. Sexual species are defined as groups of interbreeding natural populations that are reproductively isolated from other such groups. In some instances, it is possible for individuals from different species to interbreed, but the process is not reproductive — the offspring is not fertile. There are a lot of such examples in nature but the best known example is the mule which is an unfertile offspring of the donkey and the horse.

*Corresponding author.

Usually, genetic recombination is considered as very beneficial for the population evolution. It allows the population to escape the mutational meltdown, which has been shown also by computer simulations.¹ On the other hand, the recombination rate in the natural populations is highly restricted and its frequency is for example of the order of one cross-over per meiosis per human chromosome (there are thousands of genes in one chromosome).² One very important phenomenon limiting the recombination frequency could be the relatively high mutation rate accompanying the recombination events.³ But there could be other effects limiting the recombination frequency.

In our previous paper we have shown that computer simulated populations with low recombination rate in the limited environment size are characterized by higher mortality of the youngest individuals.⁴ Furthermore, if two populations were mixed in one environment and the competition between them was allowed, usually one population won in a relatively short time and the winning population showed higher mortality of the youngest. Thus, the limited recombination rate could lead to the specific strategy of population evolution. This strategy is connected with the possibility of forming linkage groups, which could survive for many generations without being disrupted. Populations evolving under lower recombination rate have relatively stable haplotypes and evolution tends to keep maximally diversified haplotypes in the genetic pool. In such a situation the probability of meeting of a higher number of loci with the defective alleles at the same diploid genome is lower. This situation resemble that described by de Oliveira,⁵ but it is connected with linkage groups rather than with single alleles.

In natural populations the problem of recombination is even more complicated. There are some groups of linked genes, which could be randomly combined from each parental haplotype during the gametogenesis — these are chromosomes. There are hot spots of recombination inside chromosomes at which recombination is much higher than at other points.^{6,7} Probability of recombination in the region between hot spots is very low, thus genes located between two neighboring hot spots are very strongly linked and “travel” through the genetic pool of population as one group. The frequency of alleles linked in one such group very often does not correspond to the frequency of prevalence of these alleles in the genetic pool, which is called linkage disequilibrium. Furthermore, there are some special mechanisms which lower the possibility of inbreeding (mating between relatives). Incidentally, a large, extremely polymorphic group of closely linked genes (Major Histocompatibility Complex — MHC), responsible for the efficiency of these mechanisms, is localized in the region where recombination is restricted.⁶

In this paper we show that lowering recombination rate to the level when the probability of separation of neighboring loci is comparable to the natural frequency leads to a specific distribution of defective genes in the genomes' populations, which could be considered as a first step towards speciation, because hybrids — organisms formed by crossbreeding of two diverged populations — usually are unfertile, their

life span is too short to produce offspring. To show this we used the Penna aging model.⁸

The most interesting feature of the populations simulated by the Penna model is differential selection pressure exerted on genes depending on the period of life in which these genes are switched on. Genes switched on before the reproduction period are under the strongest selection pressure and, as an effect, the fraction of defective alleles among them is the lowest. Genes switched on after the minimum reproduction age are under weaker selection pressure and, as a result, in the genetic pool of the population, the fraction of defective alleles is growing with the age at which they are switched on. The Penna model has been used several times to show these effects on genes expressed at different periods during the life span.⁹ In the sexually reproducing populations, the genome is represented by two bit-strings (every bit represents a gene) and the gamete is a product of cross-over between these two bit-strings (haplotypes). In the standard model there is one cross-over during each gamete production. Since the bit-strings in the model have usually much fewer loci than natural chromosomes, the probability of separation of two neighboring bits is much higher under such conditions than in natural genomes and any genetic linkage cannot be observed, which is why we have lowered the probability of cross-over between the bit-strings in our studies.

2. Model

For the standard Penna model, see its detailed description in Refs. 8–10. In our simulations the individual organism is represented by its diploid genome, consisting of two strings of 63 loci each switched on chronologically (the last locus — 64 is used to mark the haplotype *a* or *b* while simulating competition between populations *a* and *b* co-existing in one environment). Value 0 at the locus means that the allele is correct (wild), 1 — the allele is defective. All defective alleles are recessive, which means that both alleles at a given locus have to be defective to determine a deleterious phenotype. An organism can reproduce if it reaches the minimum reproduction age $R = 20$. To reproduce, the organisms form haploid gametes. During the gamete production, the two strings of the diploid parental genome exchange the homologous fragments with probability C at one randomly chosen position. We have performed our simulations with three C values — 0, 0.001 and 1. One of the two new strings (gametes) is randomly chosen and a new mutation is introduced into one randomly chosen locus.

If a wild allele is chosen for mutation — its value changes from 0 to 1, if a defective allele is chosen — it stays defective, its value remains 1. Each female at the reproduction age produces one gamete which is joined with a gamete produced by a randomly chosen male organism, also at reproduction age. A newborn organism has no genes switched on. In the first year (time step) both alleles in the first locus of a newborn are switched on. If at least one is correct, the determined phenotype is correct, i.e., all our mutations are recessive. If both alleles are mutated

the deleterious phenotype is expressed. At the second year, the alleles in the second locus are switched on and so on. If the number of expressed defects reaches a declared value $T = 3$, the organism dies. The organism dies also when it reaches the maximum age (the alleles of the last locus are switched on).

To avoid unlimited growth of the population, the Verhulst factor V is introduced: $V = 1 - N_t/N_{\max}$ where N_{\max} — the maximum population size — is often called the capacity of the environment, and N_t is the current population size. For each zygote a random number between 0 and 1 is generated and if it is greater than V , the zygote dies. In our simulations the Verhulst factor operates only at conception, meaning zygote formation, which implies that there are no random deaths in the population. This influences the structure of the genetic pool and the age structure of the population when comparing with simulations where the Verhulst factor operates randomly at all ages.¹¹

3. Results

As described in a previous paper,⁴ we produced 100 populations evolving independently from the initial populations for 100 000 time steps in the environment of $N_{\max} = 10\,000$. Individuals of the initial populations consisted of diploid genomes with each locus set for 1 (defective) with probability 0.1. Pairs of such populations (after 100 000 MCS of independent evolution) were put into one environment of doubled size ($N_{\max} = 20\,000$) and allowed to co-evolve until one of them was extinct. Then, we divided a set of 100 populations into subsets of 50 winning and 50 losing populations. There were two different modes of co-evolution: with and without crossbreeding. The co-evolution without crossbreeding mimics competition between two different species, while co-evolution with crossbreeding, under limited recombination rate mimics competition between haplotypes. As it was mentioned in the description of the model, we have used the last bit of the string to mark the haplotype. This is an appropriate method if there is no cross-over during the gamete production, or there is no crossbreeding between populations because there is no possibility of transfer of any allele or group of alleles from haplotype type a to haplotype type b or vice versa. Such transfer could happen in crossbreeding populations with cross-over.

3.1. *Hybrids viability*

Since it does not make sense to trace the haplotypes in competing populations if recombination between haplotypes in hybrids is allowed, we studied the viability of hybrids (diploid organisms produced from gametes derived from different populations) and compared it with the viability of organisms formed by joining the gametes derived from one population. To do this we took two populations — a winning one and a losing one. We have randomly chosen a female at the reproduction age, which produced as many gametes as males at the reproduction age in the population. Each of these gametes was joined with a gamete produced by

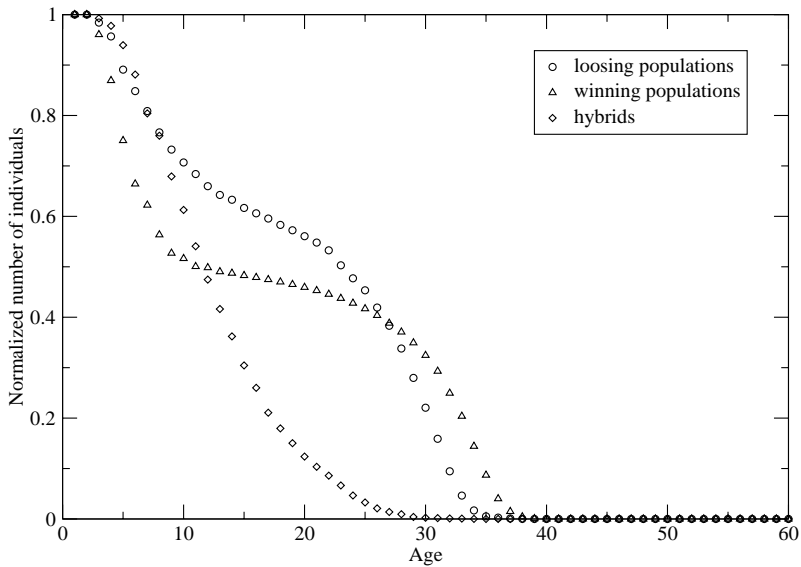


Fig. 1. Age distributions of three groups of populations. One hundred populations were simulated for 100 000 MCS without cross-over with $N_{\max} = 10\,000$. Then pairs of such populations evolved without crossbreeding in one environment with $N_{\max} = 20\,000$ until one population was extinct. Two plots represent the data averaged separately for pools of 50 winning and 50 losing populations and the third plot (\diamond) show the age distribution of hybrids — diploid individuals formed of two gametes produced by different populations.

a randomly chosen male also at the reproduction age. Then the next female was taken for gamete production until 1 000 000 organisms were produced. Then we have anticipated the lifespan of each obtained organism. In Fig. 1 we have shown the results of such an anticipated averaged age distribution and in Fig. 2 the mortality, separately for 50 winning populations (both gametes produced by parents from one winning population), for 50 losing populations (both parents from one losing population) and for hybrids formed in analyzed pairs of populations (one parent from a winning population and the other from a losing one). Note that in these results there is no effect of recombination between haplotypes — these results represent the genetic status of organisms of the first generation after crossbreeding between them was allowed. The mortality of hybrids was significantly higher at any age. This effect was also visible for populations with low cross-over rate (0.001) but disappeared for populations with high cross-over rate — 1 (results not shown).

3.2. Intraspecies competition between haplotypes

As noticed previously, when two independently evolved populations were put into one environment as two species (no crossbreeding allowed) one population won after about 2000 MCS.⁴ The winning populations are characterized by a higher mortality

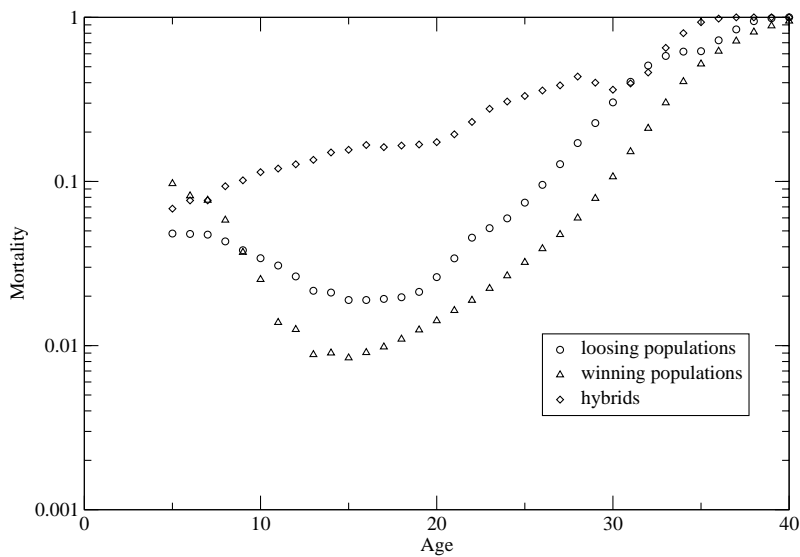


Fig. 2. Averaged mortality in sets of populations described in Fig. 1.

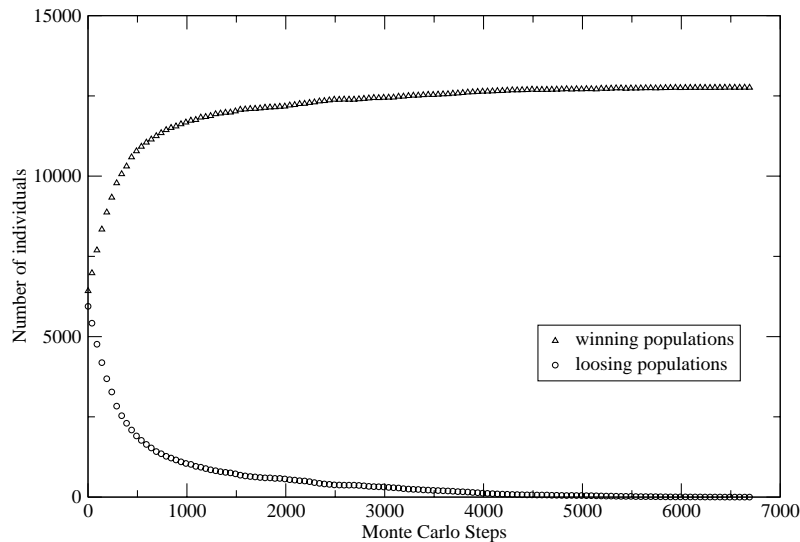


Fig. 3. The competition between two populations without crossbreeding (two species model) in the shared environment.

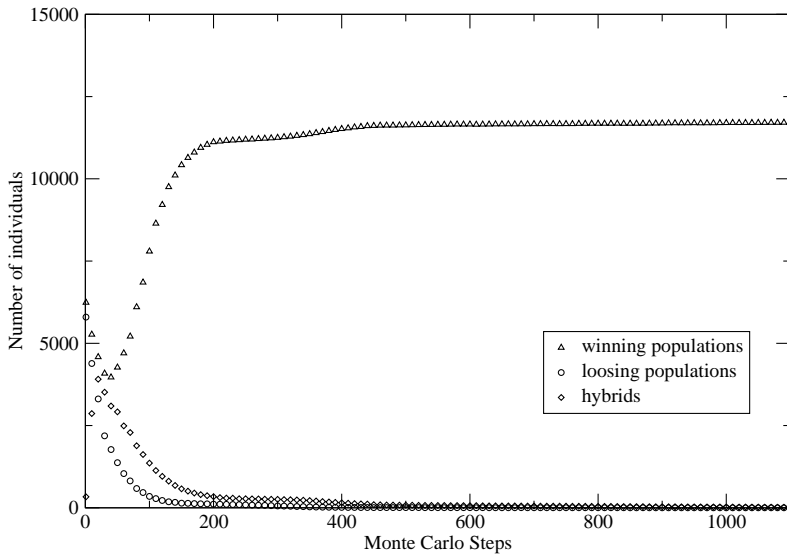


Fig. 4. The competition between two populations with crossbreeding, but without recombination between haplotypes (one species model) in the shared environment. Note the third populations — hybrids and compare the time scale in Figs. 3 and 4.

of the youngest and lower mortality and higher fraction of individuals reaching the reproduction age. Nevertheless, when crossbreeding between populations was allowed, the hybrid individuals could arise, whose genomes were constructed of two haplotypes, each derived from a different population. The lifespan of hybrids was much shorter than that of any of the two parental populations. As a result, the fraction of hybrids reaching the minimum reproduction age was much lower. This has serious consequences for the co-evolution of the two crossbreeding populations.

For populations without cross-over during the gamete production and without crossbreeding (competition of two species), the average number of time steps necessary for elimination of one population was about 2000 (Fig. 3), while it takes only 200 MCS when the crossbreeding was allowed (competition of sets of haplotypes within one species, Fig. 4).

3.3. Divergence of genetically separated populations

We have repeated the simulations in the same conditions as described above with only one difference — instead of simulating the evolution of each independently generated initial population, we have first simulated the evolution of one initial population for 10 000 MCS, then the obtained population was treated as an initial one for two further independent simulations with different random number seeds for another 100 000 MCS. In such a way we have obtained 50 pairs of populations,

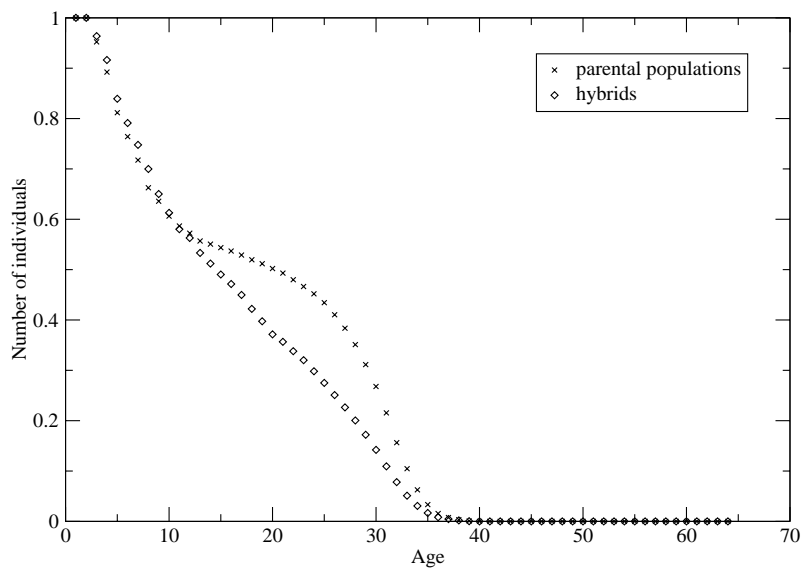


Fig. 5. The averaged age distribution of parental populations generated by independent evolution from one initial population and populations of hybrids (see text for the description of history of populations' evolution).

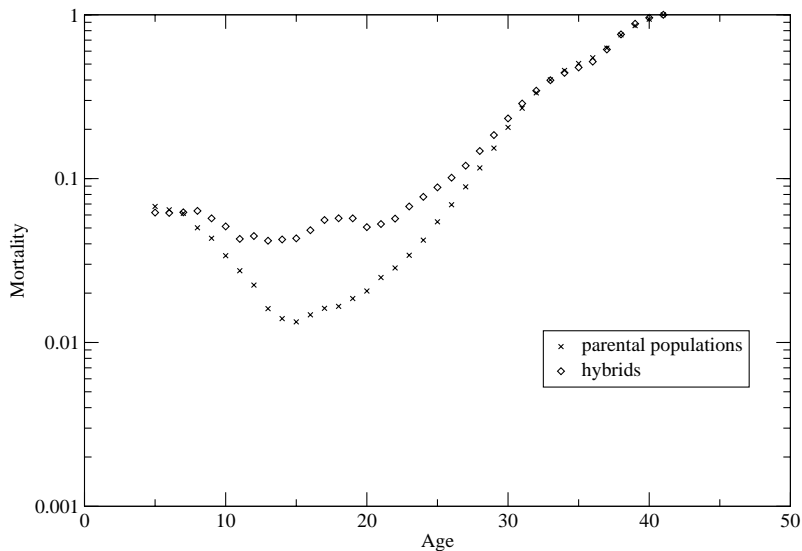


Fig. 6. Averaged mortality in sets of populations described in Fig. 5.

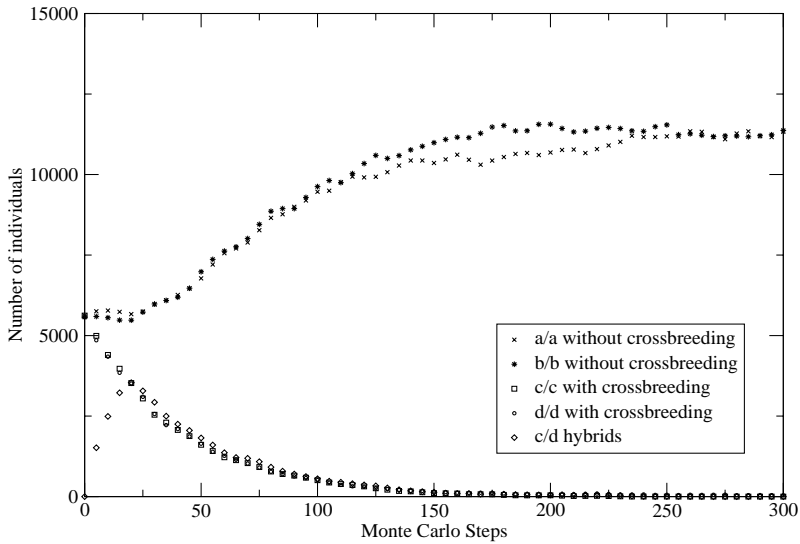


Fig. 7. The competition between two pairs of populations. Initially both pairs are identical — population *a* corresponds to population *c* and population *b* corresponds to population *d*. Simulating competition between pairs crossbreeding between *a* and *b* has been forbidden while populations *c* and *d* could freely crossbreed. Compare the time scale in Figs. 3–5.

in each pair both populations evolved from one ancestor population and diverged for another 100 000 MCS. The results are presented in Figs. 5 and 6 and are much the same as the results obtained from independently generated initial populations which indicates that genetic separation of populations for sufficiently long time leads to such changes in the genetic pool of these populations that the hybrids are less viable. The best strategy for populations to survive in one environment for a longer time is to avoid crossbreeding, which means to form a barrier limiting the crossbreeding. To show that, we put two pairs of populations *a/b* and *c/d* into one environment. Initially, the pairs were exactly the same, but the populations in the pair *a/b* were not allowed to crossbreed while the populations in the pair *c/d* could crossbreed. Thus, the populations of the pair *a/b* were different species while the populations of the second pair belong to the same species. The behavior of the populations is shown in Fig. 7. The crossbreeding populations were totally eliminated in few generations.

4. Conclusions

Many specific phenomena in the population evolution can be observed when the recombination rate between haplotypes during the gamete production is limited to the level comparable to the natural frequency of crossover. One of the most unexpected features is a higher mortality of the hybrid organisms. One could explain this effect by the strong selection against the accumulation of haplotypes with the same

distribution of defective alleles in the genetic pool. In the population with a low “biodiversity”, the frequency of homozygous recessive loci in the diploid genome would be higher and, as a result, the mortality would be higher, too. Two independently evolving populations can have “overlapping” haplotypes, which, when meet in crossbreeding populations form hybrids with more frequently occurring homozygous loci than in each of the original population, that is why the mortality of hybrids is higher. The by-product effect of this phenomena is a faster elimination of one co-evolving population when crossbreeding is allowed. In the case of two co-evolving species, competing for the same environment, the ratio between the number of newborns of $a//a$ genotype to the newborns of $b//b$ genotype equals to the ratio between the fraction of individuals in the reproduction age of both genotypes. If the two populations can crossbreed, the ratio between the number of offspring of a given genotype is described by a Newton binomial $a^2 + 2ab + b^2$. Thus, assuming that no hybrid survives until the reproduction age, the ratio between numbers of offspring of the two genotypes $a//a$ and $b//b$ roughly equals the square of fraction of individuals in the reproduction age of the two genotypes, which leads to much faster elimination of the population with a lower fraction of individuals in the reproduction age in crossbreeding populations and to the elimination of both crossbreeding populations if they compete with genetically separated populations.

Acknowledgements

The authors thank D. Stauffer for comments and discussions. The work was supported by the grant number 1016/S/IMi/02.

References

1. A. T. Bernardes, *J. Physique I* **5**, 1501 (1995).
2. P. S. White and T. C. Matise, *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 2nd edition, eds. A. D. Baxevanis and B. F. F. Oulette (Cambridge University Press, Cambridge, UK, 2001).
3. I. P. Gorlov and O. Y. Gorlova, *J. Theor. Biol.* **213**, 1 (2001).
4. A. Laszkiewicz, E. Niewczas, Sz. Szymczak, A. Kurdziel, and S. Cebrat, *Int. J. Mod. Phys. C* **13**, 967 (2002).
5. P. M. C. de Oliveira, *Theor. Biosci.* **120**, 1 (2001).
6. D. B. Goldstein, *Nat. Genet.* **29**, 109 (2001).
7. G. A. Huttley, M. W. Smith, M. Carrington, and S. J. O'Brien, *Genetics* **152**, 1711 (1999).
8. T. J. P. Penna, *J. Stat. Phys.* **78**, 1629 (1995).
9. S. M. de Oliveira, P. M. C. de Oliveira, and D. Stauffer, *Evolution, Money, War and Computers* (Teubner, Stuttgart-Leipzig, 1999).
10. D. Stauffer, *Comput. Phys.* **10**, 341 (1996).
11. J. S. Sa Martins and S. Cebrat, *Theor. Biosci.* **119**, 156 (2000).